



专题:大数据平台与应用

分布式文件系统与 MPP 数据库的混搭架构在电信大数据平台中的应用

张 雨¹, 蔡 鑫², 李爱民³, 刘 翼³, 姚晓辉²

(1. 中国电信股份有限公司上海分公司 上海 200122;

2. 中国电信股份有限公司上海研究院 上海 200122; 3. 中国电信集团公司 北京 100032)

摘 要:通过对分布式文件系统和 MPP 数据库这两种大数据技术领域重要方向的研究比较,提出了采用混搭架构进行大数据平台架构规划和设计的思路 and 实现方法,并在具体项目中进行了实际验证。该架构能够兼具两种技术的优点,适应不同场景的大数据存储和处理要求。

关键词:分布式文件系统;MPP 数据库;混搭架构

doi: 10.3969/j.issn.1000-0801.2013.11.003

Application of Distribute File System & MPP Database Mashup Architecture in Telecom Big Data Platform

Zhang Yu¹, Cai Xin², Li Aimin³, Liu Yi³, Yao Xiaohui²

(1. Shanghai Branch of China Telecom Co., Ltd., Shanghai 200122, China;

2. Shanghai Research Institute of China Telecom Co., Ltd., Shanghai 200122, China;

3. China Telecom Corporation, Beijing 100032, China)

Abstract: Distributed file system and MPP database are two important directions in the domain of big data. Based on the research of the two techniques, the sketch and implement method of big data platform architecture planning and design were put forward using the mashup architecture, and the actual verification in the specific project was carried out. This architecture can combines the advantages of both two techniques, and can adapt to the different scenes of big data storage & processing requirements.

Key words: distributed file system, MPP database, mashup architecture

1 引言

移动互联网时代,运营商普遍面临管道化的风险,客户价值受到第三方内容和应用提供商的不断挤压,国内三大运营商的战略转型迫在眉睫。中国电信集团公司董事长王晓初在 2011 年提出的“一去两化新三者”,被视为当前对运营商转型策略最集中的回答。2013 年,中国电信进一步提出,计划用 5 年的时间再造一个新中国电信,使新业

务占比超过 50%,包括流量、互联网应用和 ICP 三大主流业务。

此次运营商围绕深化信息服务的转型,突破口在进一步挖掘和显性化信息价值。目前大数据的概念火热,业界普遍认同大数据是另一座“金矿”,是重要的企业资产。汇聚企业数据、挖掘数据价值,成为运营商不约而同聚焦的重点。中国电信集团公司总部和主要省公司,以汇聚 DPI(deep packet inspection,深度分组检测)数据、构

建客户特征库、支撑流量经营为切入点,开展大数据技术和应用探索。

未来的大数据平台,既要支撑对内的数据分析应用,包括面向管理层的企业管理/经营分析,面向前端的业务开发、运营与推广分析,面向后端的网络运营与优化分析,又要支撑对外的数据开放与数据经营服务,包括数据直接租售和数据产品化。

在技术上,大数据平台需要具备支撑非结构化数据以及海量结构化数据的存储和处理能力,满足不同类型数据分析应用的并发访问和响应时间的要求。大数据项目规划和设计中,首要面对的问题便是如何利用现有技术,科学合理地构建大数据平台架构。

2 相关技术的比较

目前大数据存储技术领域有两个主要的技术阵营和研究方向:一是以Hadoop为代表的分布式文件系统(distributed file system,DFS),另一个是MPP(massive parallel processing,大规模并行处理)数据库。以下对两种技术进行简要说明和比较。

2.1 分布式文件系统

分布式文件系统是指文件系统管理的物理存储资源不一定直接连接在本地节点上,而是通过计算机网络与节点相连。分布式文件系统的设计基于客户机/服务器模式。一个典型的网络可能包括多个供多用户访问的服务器。另外,对等特性允许一些系统扮演客户机和服务器的双重角色。例如,用户可以“发表”一个允许其他客户机访问的目录,一旦被访问,这个目录对客户机来说就像使用本地驱动器一样。

Hadoop 分布式文件系统(HDFS)是运行在通用硬件上的分布式文件系统,提供了一个高度容错性和高吞吐量的海量数据存储解决方案。HDFS已经在各种大型在线服务和大型存储系统中得到广泛应用,成为各大网站等在线服务公司的海量存储事实标准,多年来为网站客户提供了可靠、高效的服务。HDFS可以提供以下特性:

- 可自我修复的分布式文件存储系统;
- 高可扩展性,无需停机动态扩容;
- 高可靠性,数据自动检测和复制;
- 高吞吐量访问,消除访问瓶颈;
- 使用低成本存储和服务器构建。

2.2 MPP 数据库

MPP由多个SMP服务器通过一定的节点互连网络进行连接,协同工作,完成相同的任务,从用户的角度看是一个服务器系统。其基本特征是由多个SMP服务器(每个SMP服务器称为一个节点)通过节点互连网络连接而成,每个节点只访问自己的本地资源(内存、存储等),是一种完全无共享(share nothing)结构,扩展能力好,目前的技术可实现512个节点互联。

在MPP系统中,每个SMP节点也可以运行自己的操作系统、数据库等。每个节点内的CPU不能访问另一个节点的内存。节点之间的信息交互是通过节点互连网络实现的,这个过程一般称为数据重分配(data redistribution)。MPP服务器需要一种复杂的机制来调度和平衡各个节点的负载和并行处理过程。目前一些基于MPP技术的服务器往往通过系统级软件(如数据库)屏蔽这种复杂性。对于MPP数据库来说,它们有以下一些特点^[1]:

- 一般存储的是结构化数据,有明显的星型或雪花型结构,适用于大数据分析的应用;
- 每个服务器都有自己独立的存储、内存和CPU,允许动态地增加或删除节点;
- 数据分区划分到不同的物理节点上,通过分布式的查询优化来提高系统整体性能;
- 主要用在数据仓库和大规模的分析处理应用中。

2.3 单一 Hadoop 技术的缺陷

Hadoop最初起源于Yahoo!公司创建的搜索引擎项目Nutch,受到Google公司GFS和MapReduce的启发,Yahoo!将Nutch改写成具有高可扩展性的分布式应用。由此可见,其设计初衷是为了实现海量非结构化数据(主要是文本)的分析处理,其支持的数据分析应用通常是批处理模式,应用需要顺序读取目标数据集的全部或大部分数据。

为了更好地存储结构化和半结构化数据以及弥补Hadoop随机读写的不足,在Hadoop之上构建了列存储的NoSQL数据库HBase,HBase具有高性能、高可靠、可伸缩的特点^[2]。尽管如此,当面对快速读写、多表关联等需求场景时,包括HBase在内的单一Hadoop技术仍显得力不从心,主要表现在以下几点。

- HBase建立在为批量分析而优化的分布式文件系统之上,天然地不适合于高速、随机访问的工作负载,这是HBase低性能的直接原因,尤其对于读取而言^[3]。



- Hadoop 的底层编程接口技术是 Java,普通用户存在使用障碍,尽管上层应用 Hive 提供了类 SQL 的 HQL 语言使用户直接对数据进行查询,Hive 负责 HQL 与 MapReduce 底层 Java 接口的“翻译”,但 HQL 并未完整地支持相关 SQL 标准,特别是对于相对复杂的多表关联查询操作显得无能为力。

3 混搭架构的设计实现

3.1 混搭架构的设计思路

根据以上分析,分布式文件系统和 MPP 数据库是大数据领域的两个技术方向。Hadoop 分布式文件系统以其水平扩展性、容错性和非结构化数据支持能力,成为大数据解决方案的首选。而 MPP 数据库同样是一种重要的大数据技术,对于快速读写、多表关联等 Hadoop 难以解决的问题,能够给出很好的答案。

经过对两种技术实现原理的深入研究、两者技术特点的客观比较、各方专家意见的广泛听取、业界应用经验的充分消化,笔者的判断是:电信大数据平台分析应用场景复杂多变,单一的技术难以支撑电信的大数据分析需求。

鉴于此,在对大数据平台技术架构进行规划和设计时,并没有生硬地做非此即彼的排他性选择,而是尝试采用分布式文件系统和 MPP 数据库的所谓“混搭架构”,实现两种技术在一个大数据平台中的共存,从而力求做到两种技术的优势互补,满足更加广泛的大数据分析应用需求。

3.2 平台技术架构的具体实现

基于分布式文件系统和 MPP 数据库混搭架构实现的大大数据平台技术架构如图 1 所示。在大数据平台的数据存储和处理层,基于统一的 x86 开放架构 IT 基础设施资源,分别搭建 Hadoop 平台和 MPP 数据库平台,二者协同配合,共同承载大数据存储和处理的任。在 Hadoop 和 MPP 数据之间,有一条单向数据链路,负责将 Hadoop 处理和规范化过的结果数据传输同步到 MPP 数据库。

大数据平台的数据源大致可以分为两块:一是来自 BOSS 的、传统的结构化数据(可称为小数据),二是 DPI、信令等海量的非结构化或半结构化的大数据。

前一部分数据相对规整,不需要进行

复杂的批量处理操作,可以直接进入 MPP 数据库,后一部分数据,则需要装载到 Hadoop 分布式文件系统,并对原始数据进行标准化处理,过滤无效数据以及将用户的行为数据与基础信息库进行匹配,其间还有可能需要对未知的网页 URL 进行内容文本的爬取和分析——MapReduce 最终处理完成的数据,是为客户行为打上分类标签的宽表数据,一方面可以直接向上提供清单级数据服务,另一方面可以持久化到 MPP 数据库,为后续融合分析提供服务。

在 MPP 数据库内,从下方数据源直接载入的 BOSS 用户资料、客户资料等数据以及水平方向 Hadoop 提供的客户行为标签数据能够在此整合,使得分散在不同网络节点上的数据得以关联,从而拼接得到完整的客户视图,获得更加深入的客户洞察,形成真正意义上的企业数据资产。

3.3 混搭架构的应用场景

根据各类数据分析应用的技术不同实现手段,可以大致将数据分析应用分为报表分析、交互分析、清单提供、数据挖掘等几种类型。

报表分析,是指利用数据可视化展现工具,实现固定报表、多维分析报表的快速开发,可发布至企业数据门户界面。市场上大多数报表工具对 Hadoop 的支持度还不够,而 MPP 数据库通过将标准的 SQL 作为数据接口,与主流报表工具能够很好地集成。因此,对于报表分析应用,选择 MPP 数据库进行支撑比较适宜。典型的报表分析应用包括业务发展分析、流量分析、终端分析、用户行为分析、产品运营分析等。

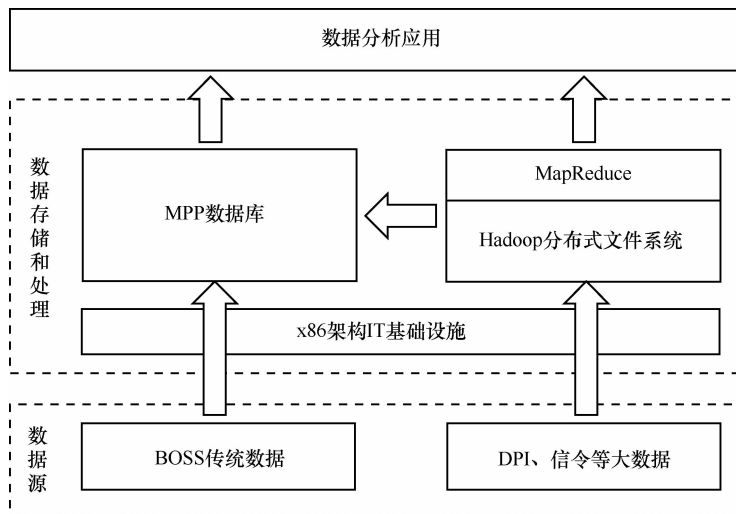


图 1 电信大数据平台之分布式文件系统与 MPP 数据库混搭架构

- 交互分析:通过配置化的操作界面,实现用户可定制的数据自助查询、自助取数和简单数据分析功能。这类应用需要根据用户的操作,生成动态的 SQL 语句,还可能涉及多表关联查询,较适宜通过 MPP 数据库进行支撑。典型的交互分析应用包括:自助取数、自助分析等。
- 清单提供:指按照固定的逻辑,周期或实时地输出清单级数据,供上层分析应用使用。此类应用的特点是批量清单级数据的顺序读取,可通过 HDFS 或 HBase 支撑。典型的清单提供应用包括:企业内部的精准营销推荐客户列单、企业外部的清单级数据租售以及数据开放服务等。
- 数据挖掘:指在海量数据中发现深层的潜在价值。用户互联网访问行为、信令位置数据大量地丰富了客户理解的数据来源,同时海量数据的计算也给传统的数据库平台带来了压力。较为一般的模式,是 Hadoop 平台负责对用户原始的行为清单数据进行清晰、转换、加工,形成标准化、标签化的行为清单数据;MPP 数据库平台在用户行为清单数据的基础上建立数据挖掘模型。当然,在 Hadoop 平台里包含了更加完整的原始数据,也可以直接在 Hadoop 分布式文件系统内进行数据挖掘,目前支持 Hadoop 的主流挖掘建模工具有 Mahout、R 等。典型的大数据挖掘应用包括:客户行为标签分类、文本分析、客户细分等。

综上所述分布式文件系统和 MPP 数据库适用的分析应用类型,可以通过表 1 概括。

表 1 分布式文件系统和 MPP 数据库适用分析应用类型对照

对比项	分布式文件系统	MPP 数据库
报表分析		√
交互分析		√
清单提供	√	
数据挖掘	√	√

4 混搭架构的现网验证

分布式文件系统和 MPP 数据库的混搭架构,率先在中国电信集团公司互联网增值业务集市平台(以下简称互增集市平台)中得到实际的现网验证。

互增集市平台主要基于基地和移动互联网业务特征

分析,提供移动互联网流量经营、终端分析等支撑。互增集市平台的技术架构说明如下。

- 平台的数据源包含结构化的基础宽表数据、基地数据、ODMS 数据、终端自注册数据以及结构化或非结构化的详单数据和网络日志数据。
- 平台的数据存储和处理层,基于云架构的 IT 基础设施,采用海量数据并行处理架构的 MPP 数据库与分布式文件存储/计算相结合的数据存储和处理技术。
- 应用层主要包括以图表等为呈现方式的交互式分析应用和以面向基地和增值业务营销的客户视图数据服务应用。其中,Hadoop 分布式文件系统主要提供基于海量清单级的、可变维度的和用户级的数据分析和挖掘服务,典型的应用场景如基础标签库应用、基地业务精准营销清单推送等;MPP 数据库主要提供相对固定维度和关联关系的数据查询和分析服务以及高速数据提取服务,典型的应用场景有:基地业务运营分析、互联网增值业务用户行为分析等。

通过互增集市平台的现网验证,说明分布式文件系统和 MPP 数据库的集成具有可行性,混搭架构能够取两种技术之所长,适应更广泛的数据应用场景需求,平台的功能和性能能够达到业务目标要求,分布式文件系统和 MPP 数据库的混搭架构在电信大数据平台建设中具有实施推广价值。

5 结束语

应当说,大数据技术的演进之路从未止步,分布式文件系统和 MPP 数据库两个阵营也都在持续推动着技术向更全面、更完备的方向发展。未来,两种技术有走向融合的迹象。

目前已经出现了在这方面进行尝试的商业产品,例如:Teradata 的 Aster Data nCluster 将 SQL、MPP 和 MapReduce 进行了整合,产品中包含一项专利技术称作 SQL-MapReduce,该项技术是由能够处理结构化数据的常规数据仓储技术和处理非结构化数据的大数据处理技术混合而成。同时 Teradata 还宣布与 Hortonworks 战略合作,推进 Hadoop 与 Teradata MPP 数据库的融合^[4]。

此外,EMC Greenplum 宣布了一个新的 Hadoop 发行版本——Pivotal HD,其主要组件是 HAWQ,这是一个 MPP 关



系数据库,借助一种动态流水线机制直接运行于 HDFS 之上。EMC 宣称,Greenplum Pivotal HD 结合了 SQL 和 Hadoop 的优势,而且速度“比 Hive 快数百倍”^[5]。

或许,在不久的将来,无需再集成不同的产品以实现分布式文件系统与 MPP 数据库的混搭,而是仅仅通过一个融合化的商业产品就能够实现两种技术的完美混搭。

参考文献

- 1 周宝曜,刘伟,范承工.大数据:战略、技术、实践.北京:电子工业出版社,2013
- 2 Apache HBase. <http://hbase.apache.org/>
- 3 Big data debate: will HBase dominate NoSQL. <http://www.informationweek.com/software/enterprise-applications/big-data-debate-will-hbase-dominate-nosql/240159475>
- 4 MapReduce and MPP: two sides of the big data coin. <http://www.zdnet.com/blog/big-datamapreduce-and-mpp-two-sides-of-the-big-data-coin/121>
- 5 Greenplum Pivotal HD 结合了 SQL 和 Hadoop 的优势. <http://www.infoq.com/cn/news/2013/03/Pivotal-HD-SQL-Hadoop>

(上接第 11 页)

- 10 Jain A, Zongker D. Feature selection: evaluation, application and small sample performance. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(2):153~158
- 11 Gorvan A. Principal Manifolds for Data Visualisation and Dimension Reduction. New York: Springer, 2007
- 12 Verleysen M, Lee J A. Rank-based quality assessment of nonlinear dimensionality reduction. Proceedings of the 16th European Symposium on Artificial Neural Networks, Bruges, Belgium, 2008:49~54
- 13 Deerwester S. Indexing by latent semantic analysis. Journal of American Society for Information Science, 1990, 41(6):391~407
- 14 陈燃燃. 基于 SVM 算法的 Web 分类研究与实现. 北京邮电大学硕士学位论文, 2009
- 15 Vapnik V N. The Nature of Statistical Learning Theory. New York: Springer, 1995
- 16 Joachims T. Text categorization with support vector machines: learning with many relevant features. Proceedings of the 10th European Conference on Machine Learning, Chemnitz, Germany, 1998:137~142
- 17 Joachims T. Transductive inference for text classification using support vector machines. Proceedings of the 16th International

[作者简介]



张雨,男,中国电信股份有限公司上海分公司高级工程师、副总工程师,主要研究方向为 IT 规划、大数据。

蔡鑫,男,中国电信股份有限公司上海研究院高级工程师、数据分析师,主要研究方向为数据规划、信息管理、数据标准。

李爱民,女,中国电信集团公司企业信息化事业部工程师、数据处处长,主要研究方向为信息管理。

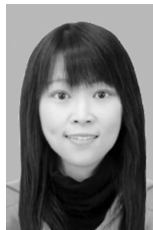
刘翼,男,中国电信集团公司企业信息化事业部工程师、业务主管,主要研究方向为数据规划、信息管理。

姚晓辉,男,中国电信股份有限公司上海研究院工程师,主要研究方向为数据规划、数据挖掘、信息管理。

(收稿日期:2013-10-25)

Conference on Machine Learning, Bled, Slovenia, 1999: 200~209

[作者简介]



边凌燕,女,中国电信股份有限公司上海研究院工程师,主要研究方向为数据分析、数据挖掘。



贺仁龙,男,中国电信股份有限公司上海研究院工程师,主要研究方向为信息管理。

姚晓辉,男,中国电信股份有限公司上海研究院工程师,主要研究方向为数据规划、数据挖掘、信息管理。

(收稿日期:2013-10-29)